AD-A240 726

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1

...ion is estimated to average 1 hour per response, including the time for reviewing instructions, search... ...g the data needed, and completing and reviewing the collection of information. Send comments re... ...f this collection of information, including suggestions for reducing this burden, to Washington Head... ...itions and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the

1. AGENCY USE ONLY *(Leave blank)*	2. REPORT DATE July 1991	3. REPORT TYPE AND DATE COVERED Journal Article 1991

4. TITLE AND SUBTITLE Dynamic Administration of a General Intelligence Test	5. FUNDING NUMBERS None
6. AUTHOR(S) G. E. Larson, D. L. Alderton, M. A. Kaupp	
7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Navy Personnel Research and Development Center San Diego, California 92152-6800	8. PERFORMING ORGANIZATION JA-91-08
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)	10. SPONSORING/MONITORING Journal of Learning and Individual Differences, Vol. 3, pp. 123-134, 1991
11. SUPPLEMENTARY NOTES	
12a. DISTRIBUTION/AVAILABILITY STATEMENT Approved for public release; distribution is unlimited.	12b. DISTRIBUTION CODE A

13. ABSTRACT *(Maximum 200 words)*

The benefits of dynamic testing are thought to include; (a) reduction in strategic variance, accompanied by (b) a test score increase for "disadvantages" subjects. Sometimes forgotten, however, is that these accomplishments are illusory unless they support a specified goal (e.g., better validity). In the present study, we examine the benefits of dynamic test administration with the Raven's Advance Progressive Matrices (APM) test of general intelligence. The results indicate that, while APM scores were significantly increased by dynamic procedures, important criteria such as reliability and construct validity were not enhanced. We conclude that the choice of dynamic procedures depends on both the ability to construct and the purpose of testing, and should be justified on a case-by-case basis.

91-11462

14. SUBJECT TERMS dynamic testing; intelligence; validity; Armed Services Vocational Aptitude Battery; ASVAB; Raven's Matrices Test; training.	15. NUMBER OF PAGES 12
	16. PRICE CODE

17. SECURITY CLASSIFICATION OF REPORT	18. SECURITY CLASSIFICATION OF THIS PAGE	19. SECURITY CLASSIFICATION OF ABSTRACT	20. LIMITATION OF ABSTRACT
UNCLASSIFIED	UNCLASSIFIED	UNCLASSIFIED	UNLIMITED

NSN 7540-01-280-5500

9 1 9 2

Standard Form 298 (Rev. 2-89)
Prescribed by ANSI Std. Z39-18
298-102

DYNAMIC ADMINISTRATION OF A GENERAL INTELLIGENCE TEST

GERALD E. LARSON, DAVID L. ALDERTON, AND MARK A. KAUPP

NAVY PERSONNEL RESEARCH AND DEVELOPMENT CENTER

ABSTRACT: The benefits of dynamic testing are thought to include; (a) a reduction in strategic variance, accompanied by (b) a test score increase for "disadvantaged" subjects. Sometimes forgotten, however, is that these accomplishments are illusory unless they support a specified goal (e.g., better validity). In the present study, we examine the benefits of dynamic test administration with the Raven's Advanced Progressive Matrices (APM) test of general intelligence. The results indicate that, while APM scores were significantly increased by dynamic procedures, important criteria such as reliability and construct validity were not enhanced. We conclude that the choice of dynamic procedures depends on both the ability construct and the purpose of testing, and should be justified on a case-by-case basis.

In recent years there has been growing interest in alternative testing procedures such as "dynamic assessment" (Feuerstein 1979) and "testing-the-limits" (Carlson & Weidl 1979), partly because of suggestions that these procedures might be superior to conventional psychometric test methods. For example, one problem with conventional testing is that identical test scores can mean different things for different people, as in the following illustration. Suppose that two individuals are taking a test of spatial ability. If Person A solves the items through transformations of mental images, while Person B uses an entirely verbal strategy, then their test scores do not reflect the same underlying construct. Hence, the construct validity of the test is impeached, and predictive validity may suffer as well. Dynamic testing may solve this problem because its procedures (including directed practice) can encourage subjects to exhibit the knowledge and skills that the test was designed to measure, and to abandon irrelevant strategies. The expected result would be a more construct-valid test. This, indeed, was the result of a study by Embretson (1987), who compared the performance of subjects tested on the figure-folding task from the Differential Aptitude Test (DAT), under either dynamic or control conditions. Embretson's

Direct all correspondence to: Gerald E. Larson, Testing Systems Department, Navy Personnel Research and Development Center, San Diego, CA 92152.

Learning and Individual Differences, Volume 3, Number 2, 1991 pages 123–134

Copyright © 1991 by JAI Press, Inc.
All rights of reproduction in any form reserved

ISSN 1041-6080

results show that dynamic testing (involving cues and solution modeling) led to improvements in both construct validity and predictive validity.

Hypothetically, a reduction in inter-individual strategy variance (through dynamic procedures) might also produce scores that are fairer to "disadvantaged" subjects, including examinees with little test-taking experience and/or test sophistication. Through directed practice, such subjects can be encouraged to abandon self-defeating strategies and thereby reveal their true competence. Those who believe that dynamic procedures enhance test fairness might point to studies such as Dillon and Carlson (1978), who found that ethnic group differences on reasoning tasks were narrower in a dynamic condition than in a control condition.

What if inter-individual strategies are important? Partly because of the perceived benefits mentioned above, the dynamic testing literature has thus far been almost uniformly positive. Perhaps it is time, therefore, to note that dynamic testing may not be a panacea, and that it may in fact be theoretically inappropriate for tests of general intelligence (or *g*), such as Raven's Advanced Progressive Matrices (APM) Test.[1] The general intelligence dimension is problematic because it may be *associated* with inter-individual strategic variation. Therefore, it is conceivable that strategy "standardization" through dynamic testing procedures may simultaneously diminish the usefulness of a test such as the APM, if strategies are themselves indices of *g*. And, in fact, there is evidence with which to pursue such an argument.

Haygood and Johnson (1983), for example, provide an interesting example of the greater strategic flexibility of "high-*g*" subjects given novel tasks. Haygood and Johnson employed the Sternberg (1966) memory-search task, in which subjects are asked to memorize a set of single digits (0–9) called the positive set. Next, subjects are asked whether the positive set includes a series of individually presented test digits. Performance is measured via reaction time (RT) to test items. A seldom emphasized aspect of the task is that as more digits are added to the positive set, fewer remain in the out-group or negative set. As the ratio shifts, the advantage of switching focus to the negative set increases because there are relatively fewer digits to work with. For example, a subject can verify that an item is a member of a small negative set faster than he or she can determine that it is a member of a much larger positive set—yet both methods can produce the correct answer. Of interest is Haygood and Johnson's finding that subjects who scored high on Raven's Progressive Matrices were also quicker to shift to a negative set focus and thereby take advantage of the difference in set sizes.

A similiar finding is reported in a study by Ippel and Beem (1987), who found a correlation between Raven's Matrices and the point at which subjects shifted from a clockwise to a counterclockwise direction on a mental rotation task. The shift pattern of subjects scoring high on the Raven was more "rational," in that they tended to rotate objects in the shortest direction. The use of dynamic testing procedures to reduce strategic variation might therefore work *against* the validity of *g*-measures such as Raven's Advanced Progressive Matrices (APM), if tests like the APM measure intelligence because they are strategy-ambiguous and thereby require flexibility and invention (see discussion by Kirby and Lawson, 1983).

We raise these issues not because of general misgivings about dynamic testing.

Rather, we are concerned that all procedures have limits, yet the limits of dynamic testing have thus far not been addressed. This hinders an informed decision by the testing professional who wishes to tailor his/her method to a specific situation.

PURPOSE

In the present experiment we explore the effect of dynamic testing procedures on the construct validity of the APM test. The dynamic testing package itself was designed to discourage various counter productive test-taking strategies which are known to be used by some low scoring subjects. If dynamic testing is the best way to measure general intelligence, then our interventions should produce a more valid APM score.

METHOD

SUBJECTS

Subjects were 808 male Navy recruits (mean age 19.5 years, SD = 2.36 years) selected at random from groups undergoing in-processing at the Recruit Training Command, San Diego.

PROCEDURE

1. Armed Services Vocational Aptitude Battery (ASVAB) scores were gathered from the recruits' personnel records. The ASVAB is a set of ten tests (listed in Table 1) used for selection and classification of military applicants. The tests are scaled to a mean of 50 and a standard deviation of 10 in an unselected, nationally representative sample.
2. Raven Progressive Matrices. Subjects were group-tested on the 36 item Raven Progressive Matrices (APM), Advanced Set II (Raven 1962), under either "standard" (N = 413) or dynamic testing conditions (N = 395) as described below.[2] One group of about 40 subjects was tested per day. Group assignment was random. In both the dynamic testing and control conditions, a time limit was set for solving each item. Time limits were necessary because correct answers were presented as part of dynamic testing, making it necessary to have subjects respond before hearing the presentation. Proctors insured that all subjects responded when told to do so.

The following item time limits were used, based on several days of pilot testing during which proctors determined the typical item solution latencies.

Items 1 to 12 : 30 seconds per item
Items 13 to 20 : 1.0 minute per item
Items 21 to 26 : 2.0 minutes per item
Items 27 to 36 : 2.5 minutes per item

The total item solution time was thus 51 minutes, which is about 25% longer than the 40 minute test-time limit suggested in the Raven manual.

Condition 1: Standard Procedures. In this condition, subjects were first given a four-item practice booklet with the following printed instructions:

> You will be shown a series of items in which part of a pattern is missing. You must pick the missing part from 8 possible choices. Be sure and look both down and across the pattern before making your choice. PLEASE CIRCLE THE CORRECT ANSWER.

The practice items were relatively easy problems taken from the APM, Set I. The test administrator read the instructions aloud. and then subjects were given as much time as needed to complete the booklet. When all subjects had finished the proctor gave out the correct answers, along with a very brief explanation of why each answer was correct. Subjects were allowed to request clarification at any time during these explanations.

Next, subjects were given the 36 item APM Set II test, with the time limits for items noted above. At the end of the time limit for each item, subjects entered their response on the numeric keypad of a Hewlett-Packard Integral Personal Computer.

TABLE 1
Tests in ASVAB Forms 11,12, and 13

Tests	*Abbreviation*	*Description*
General Science	GS	A 25-item test of knowledge of the physical (13 items) and biological (12 items) sciences—11 minutes
Arithmetic Reasoning	AR	A 30-item test of ability to solve arithemtic word problems —36 minutes
Word Knowledge	WK	A 35-item test of knowledge of vocabulary, using words embedded in sentences (11 items) and synonyms (24 items) —11 minutes
Paragraph Comprehension	PC	A 15-item test of reading comprehension—13 minutes
Numerical Operations	NO	A 50-item speeded test of ability to add, subtract, multiply, and divide one- and two-digit numbers—3 minutes
Coding Speed	CS	An 84-item speeded test of ability to recognize numbers associated with words from a table—7 minutes
Auto and Shop Information	AS	A 25-item test of knowledge of automobiles, shop practices, and use of tools—11 minutes
Mathematics Knowledge	MK	A 25-item test of knowledge of algebra, geometry, fractions, decimals, and exponents—24 minutes
Mechanical Comprehension	MC	A 25-item test of knowledge of mechanical and physical principles—19 minutes
Electronics Information	EI	A 20-item test of knowledge of electronics, radio, and electrical principles and information—9 minutes

Condition 2: Dynamic Testing. The purpose of the APM dynamic testing package was to reduce self-defeating and/or irrelevant APM strategies, by using performance-enhancing methods from previous studies. Two of these self-defeating strategies deserve special mention. They are (a) gestalt (or imagery-based) algorithms and (b) abbreviated encoding. Regarding the former, Hunt (1974) suggested that at least two algorithms could be used to solve APM items. One, called the *gestalt algorithm*, emphasizes the operations of visual perception, such as the continuation of lines through blank areas and the superimposition of visual images upon each other. For example, one might imagine the existing lines in the matrix stretching into the missing cell, without thinking abstractly about the properties of the matrix. The second algorithm, called the *analytic algorithm*, breaks the matrix elements down into features, then employs logical operations to determine which features and relationships are critical. Various analyses suggest that the gestalt algorithm is inferior because it cannot be used to solve difficult items (e.g., Hunt 1974; Kirby & Lawson 1983; Lawson & Kirby 1981). Therefore, our dynamic procedures discouraged its use.

The second self-defeating strategy (or tendency) we attempted to discourage involves abbreviated problem encoding, sometimes associated with impulsivity. Lawry, Welsh, and Jeffrey (1983), for example, categorized children as reflective or impulsive (using Kagan's Matching Familiar Figures test) and then studied the performance of these two groups on matrix items. Lawry et al. found that as the items became more difficult, reflectives slowed their performance more than did impulsives. Moreover, slowed responding was associated with higher scores on the more difficult items, a finding later replicated by Welsh (1987). Correction of this tendency is therefore a proper goal of dynamic testing, particularly since impulsivity is thought to disproportionately impair the performance of lower socioeconomic class subjects (see Turner, Hall, & Grimmett 1973).[3]

To summarize the rationale behind the design of the dynamic testing package, we sought to encourage subjects to use optimal analytic strategies. Moreover, proponents of dynamic testing might argue that our interventions should improve validity. The specific dynamic procedures are described below.

Part 1: Analytic Reflective Instructions. Part 1 involved a test booklet (12 items, 12 minutes) designed to encourage subjects to (a) adopt an analytic approach to inferring the rule underlying the matrix pattern, and to (b) generate and sketch their own item solutions *before* looking at the answer choices. The items were all from the APM Advanced Set I.

The cover page of the booklet offered the following advice:

1. Remember that all problems can be broken down into smaller steps. The steps can be worked on one at a time.
2. Make sure that your steps follow a logical order toward solving the problem.
3. Work carefully! Many mistakes are made just because people are in too much of a hurry.
4. If it pays to start over, then start over. Don't stick with something that doesn't seem to be working.

Each item in the booklet spanned 3 pages. On the first page, the matrix was presented, along with a space for drawing a picture that would complete the pattern. The printed instructions asked the subjects to (1) analyze the changes in the matrix pattern from left-to-right and top-to-bottom, (2) figure out the rule that explains the changes, (3) use the rule to draw an answer, and (4) go on to the second page. On page 2 of the problem, subjects were again shown the matrix problem, but this time the answer choices were displayed along with it. The printed instructions asked the subjects to find and circle the answer they had drawn on page 1. If their answer was not among the choices, subjects were instructed to (1) analyze the puzzle again, (2) figure out the rule that explains the changes from left-to-right and top-to-bottom, (3) use the rule to choose an answer. They were instructed not to go back and change their drawing, but rather to go to page 3 where the *correct* answer was shown. On page 3, the matrix and the answer choices were again shown, with the correct answer circled. The design of the PART 1 booklet was influenced by analytic training methods reported in Kirby & Lawson (1983); Lawson & Kirby (1981); Malloy, Mitchell & Gordon (1987); Sternberg (1986).

Part 2: Rule Combination Principle. In part 2 of the dynamic testing session, the proctor presented examples of 12 simple item progression (or relation) rules that are common in figural analogy problems (e.g., "change in size," "change in shape"). The rules themselves, which are shown in Table 2, are adapted from Jacobs and Vandeventer (1972). For each of the 12 rules, (1) an example was presented by the proctor (using an overhead projector and a portable projection screen), following which (2) the subjects were asked to solve a second example in a booklet. Then, (3) the proctor presented the solution to the second example, again using the overhead projector. After all 12 rules had been demonstrated by the proctor and attempted by the subjects, 3 examples of "rules in combination" were presented to show how seemingly complex problems are sometimes merely combinations of several simple rules.

Part 3: Modeling. The third part of the dynamic testing procedure was embedded in the actual test session. First, subjects used the alloted time to solve each of the 36 APM Set II items (with the previously described item time limits), after which the proctors ensured that all subjects entered and recorded their answer on the computer. This was to prevent subjects from changing an answer once time had expired. After each answer was recorded, a proctor used the overhead projector to demonstrate how the problem could have been solved to obtain the correct answer. The problem solutions were read from a script. Our goal was to *model* successful, rule governed problem solving throughout the course of the test.

To summarize the dynamic testing package, Part 1 was designed to encourage subjects to view the matrix problems analytically, and to avoid impulsive answer choice selection by generating and drawing answers before examining the alternatives. Part 2 was designed to reinforce the concept that problems are rule-governed and that seemingly complex problems can (sometimes) be analyzed in terms of combinations of simpler rules, some of which were presented as part of the instruction. Part 3 allowed the proctor to model, on a continuing basis, a successful method of problem solving. A dynamic test session typically required 2.5 to 3 hours.

TABLE 2
Twelve Item Rules[a]

1. *Identical pattern:* Every cell is exactly the same.
2. *Shading:* Progressive change in shading.
3. *Movement in a plane:* Figure moves as if slid along a surface.
4. *Reversal:* Two elements exchange some feature, such as size, shading, or position.
5. *Addition:* The figure in one column (row) is added to that in the second, and the result is placed in the third.
6. *Number series:* Constant increase in items across cells.
7. *Shape:* Complete change of form, or systematic change, as from solid to dotted lines.
8. *Size:* Proportionate change, as in photographic enlargement or reduction.
9. *Mirror image:* Figure moves as if lifted up and replaced face down.
10. *Added element:* A new element is introduced, or an old one is taken away.
11. *Unique addition:* Unique elements are treated differently from common elements, e.g., they are added while common elements cancel each other out.
12. *Three of a kind:* Each element appears three times in a 3 × 3 matrix.

[a]Adapted from Jacobs & Vandeventer (1972).

RESULTS

First, we calculated a *g*-score for each individual based on the ten ASVAB tests, to use as an external (or criterion) measure of *g* when examining the construct validity of the Raven. The ASVAB-*g*, score was derived by performing a hierarchical factor analysis (orthogonalized following Schmid and Leiman 1957) on ASVAB scores from the 1988 fiscal year Navy applicant sample (N = 147,287). The loadings of the 10 ASVAB tests on the hierarchical factor were subsequently used as weights to calculate an ASVAB-g for each individual.

Next, comparison statistics were computed to contrast the "dynamic" (N = 395) vs. "control" (N = 413) groups. The results are shown in Table 3, where it can be seen that the two groups are almost identical on ASVAB-*g* scores from pre-enlistment test sessions. This further indicates that group assignment was indeed random. However, the APM scores differ significantly ($t = -8.11, p < .001$), with the dynamic testing subjects obtaining a higher mean score. This finding is not surprising since dynamic procedures are designed to be helpful. The essential question is whether a "better" test emerges, e.g., is there any gain in construct validity?

TABLE 3
Comparison of Standard and Dynamic Testing

		RAVEN			*ASVAG-g*		*APM × g*
	N	*Mean*	*SD*	*Alpha*	*Mean*	*SD*	*Correlation*
Control	413	19.39	6.19	.89	326.45	29.48	.59
Dynamic	395	22.60	5.06	.77	325.18	30.26	.57
Archival	1693	17.31	5.76	.83	324.48	30.44	.56

TABLE 4
Raven (APM) Scores and Validity as a Function of Aptitude Group and Test Condition

LOW-*g*	HIGH-*g*
CONTROL	
(N = 102)	*(N = 103)*
APM: $\bar{X}$,SD = 15.06, 5.16	APM: $\bar{X}$,SD = 24.55, 4.76
ASVAB-*g*: $\bar{X}$,SD = 288.54, 11.49	ASVAB-*g*: $\bar{X}$,SD = 364.58, 12.46
APM × *g* correlation = .27, $p < .01$	APM × *g* correlation = .34, $p < .001$
DYNAMIC TESTING	
(N = 113)	*(N = 108)*
APM: $\bar{X}$,SD = 19.35, 4.49	APM: $\bar{X}$,SD = 26.43, 4.00
ASVAB-*g*: $\bar{X}$,SD = 289.02, 12.38	ASVAB-*g*: $\bar{X}$,SD = 363.18, 11.43
APM × *g* correlation = .27, $p < .01$	APM × *g* correlation = .27, $p < .01$
ARCHIVAL	
(N = 424)	*(N = 429)*
APM: $\bar{X}$,SD = 13.58, 4.98	APM: $\bar{X}$,SD = 22.06, 4.47
ASVAB-*g*: $\bar{X}$,SD = 285.19, 11.46	ASVAB-*g*: $\bar{X}$,SD = 364.23, 12.18
APM × *g* correlation = .27, $p < .01$	APM × *g* correlation = .24, $p < .01$

To determine whether the dynamic testing APM score would be an improved measure of *g* correlations between APM and the ASVAB-*g* score were computed. The correlation between APM and ASVAB-*g* was .59 ($p < .0001$) for the "control" group, and .57 ($p < .0001$) for the "dynamic testing" group. The difference in correlations was not statistically significant. Finally, the internal consistency of the APM was assessed via Cronbach's Alpha, yielding reliabilities of .84 and .77 for scores obtained under standard and dynamic procedures, respectively. These values are significantly different ($F[412,394] = 1.438$, $p < .01$). Nothing in these results suggests that dynamic procedures enhanced the precision of the APM test.

Also shown in Table 3 is "archival" data from some of our various research studies of the last three years. The subjects were comparable to those in the present study (i.e., Navy men of about the same age). The scores were obtained with 40 minute self-paced test sessions, however, rather than the current group-paced procedure wherein time limits were set *per item*. The archival data was thus collected under procedures specified in the APM manual, without any experimental intervention whatsoever. It can be seen that the failure of dynamic procedures to improve reliability or construct validity is not a function of our control group, since a comparison of dynamic testing data with archival data leads to similiar conclusions.

In the overview we noted that the construct-irrelevant strategies targeted by our dynamic testing package are most often exhibited by "lower ability" subjects. Therefore, the greatest treatment effect (and benefits) should be found in the lower ranges.

To examine whether this was indeed the case, we constructed two subsamples, comprised of subjects in the upper and lower quartiles of the full sample distribution of ASVAB-*g* scores. Treatment effects were then examined separately for the high-*g* and low-*g* groups. Table 4 displays the within-group correlations between the APM and the ASVAB-*g*. (Recall that these correlations reflect construct validity). As the table shows, dynamic testing did not affect construct validity in the low ability subjects (.27 in both conditions). Rather, if any validity change did occur, it was for high ability subjects and in the wrong direction (.34 vs .27).

TABLE 5
Rules Governing the APM Items, and Treatment Effects for Items

Raven	*Rules*[a]	*Effect of Dynamic Training* ($p < .05$)
1	6,12	−
2	6	−
3	3	−
4	2	+
5	2	+
6	6	NS
7	5	+
8	12	NS
9	5	+
10	3,8	+
11	5	+
12	11	+
13	6,12	NS
14	3	NS
15	5	NS
16	11	+
17	12	+
18	?	NS
19	5	+
20	5,?	+
21	2,3,8,12	+
22	11	+
23	11	+
24	2	+
25	2,5	+
26	3,12	NS
27	10,12	NS
28	6,12	NS
29	3,8,12	+
30	?	NS
31	3,12	NS
32	3	NS
33	5,11	NS
34	6,12	−
35	5,11	NS
36	5,11	NS

[a]From Table 2.

Also, it must be noted that the untrained "high *g*" group in Table 4 outperformed the fully trained "low *g*" group (t [1,214] = −8.26, $p < .001$), suggesting that the APM is measuring a genuine intellectual trait on which relatively stable individual differences exist. The archival data supports these conclusions.

Finally, we constructed Table 5 to show how the 12 rules presented in training apply to the 36 APM items, and also to clarify which items were significantly enhanced by the dynamic procedures. (Item/rule relationships were independently determined by two of the authors, with a third author resolving disputes). Inspection of the table shows that, as expected, multiple rules are more common on the later (harder) APM test items. All significant treatment effects in Table 5 indicate that the dynamic testing group had a higher proportion correct than control subjects, except for the first three items, where dynamic testing was associated with a significant drop in performance. From examining the table, it is apparent that most of the score gain from dynamic procedures came in the middle portion of the test.

DISCUSSION

While dynamic testing is usually seen as supporting the construct validity of tests, it's usefulness may have limits, as in the case of a general intelligence test. One reason is that previous studies (Haygood & Johnson 1983; Ippel & Beem 1987) link *g* to the kind of strategic variation that might be diminished by dynamic procedures. Conceivably, dynamic testing might actually reduce validity. Our results, however, show little harm from dynamic administration of Raven's Advanced Progressive Matrices. Since no *psychometric benefit* was obtained either, there seems little reason to undertake time-consuming dynamic procedures for a test like the APM.

The significant raw score gain following training highlights the importance of testing all examinees under the same conditions when scores will be used operationally. However, despite significant score gains following training, evidence for the *stability* of general intelligence scores was also obtained. That is, the *untrained* "high ASVAB-*g*" group *outperformed* the fully-trained "low ASVAB-*g*" group on a second measure of *g* (the APM). Even several hours of training and problem solving demonstrations were insufficient to allow low-*g* subjects to perform like high-*g* subjects. This suggests that the Raven scores reflect somewhat unmalleable qualities such as induction and working memory capacity (see Carpenter, Just, & Shell 1990) rather than strategic differences. But if so, then why do Raven scores correlate with strategic differences on *other* tasks like memory scanning and mental rotation? One possible explanation is that if the Raven is, indeed, measuring something like working memory capacity, then subjects with high capacity may have greater "reserves" or spare capacity for self-monitoring on laboratory tasks, allowing the discovery of efficient strategies. This theory would predict that strategic differences would not be so apparent if all subjects were working at maximum capacity.

One possible criticism of the present study is that the ASVAB is not necessarily a good measure of *g* and that it is therefore an inadequate criterion by which to judge the construct validity of the Raven. In response, we turn to conventional wisdom. Convention suggests that when a 10 test aptitude battery with fair content diversity (the ASVAB) is subjected to a hierarchical factor analysis, the first hierarchical factor is a satisfactory estimate of *g*. We feel that the relationship between this hierarchical *g* and the Raven is a fair and reasonable index of the degree to which the Raven measures general intelligence. We interpret *changes* in this relationship following training as changes in the effectiveness of the Raven.

We do not see our results as a general indictment of dynamic procedures. Rather, we are primarily concerned that if the limits of these procedures must be addressed. To simply "boost" a score through dynamic procedures serves no useful purpose unless the result is more reliable and/or valid.

ACKNOWLEDGMENT: The opinions expressed in this article are those of the authors, are not official and do not necessarily reflect the views of the Navy Department.

NOTES

1. When test data are subjected to either hierarchical factoring or multidimensional scaling, reasoning tests such as Raven's Progressive Matrices are the best markers for the general variance in the battery (Marshalek, Lohman, & Snow 1983). Also, LISREL analyses indicate that reasoning tests are excellent measures of g (Undheim & Gustafsson, 1987).

2. Considerably smaller numbers of subjects underwent "partial dynamic testing." Since the N's are smaller and the results are consistent with those already being reported, there is little to be gained by presenting the partial treatment data.

3. While these studies all involved children, it is quite likely that a dimension like impulsivity affects the "encoding time" parameter in performance models on adult subjects. For example, Sternberg (1977) found that adult subjects who are relatively poor at solving analogies also tend to spend less time encoding the items. Indeed, a "deep encoding" style can be said to be a characteristic of highly skilled (or expert) performance in general (Nickerson 1988).

REFERENCES

Carlson, J.S., & K.H. Weidl. (1979). "Toward a differential testing approach: Testing-the-limits employing the Raven's matrices." *Intelligence*, 3, 323–344.

Carpenter, P.A., M.A. Just, & P. Shell. (1990). *What one intelligence test measures: A theoretical account of the processing in the Raven Progressive Matrices Test*. Pittsburgh, PA: Department of Psychology, Carnegie Mellon University.

Dillon, R., & J.S. Carlson. (1978). Testing for competence in three ethnic groups. *Educational and Psychological Measurement*, 38, 437–443.

Embretson, S.E. (1987). "Improving the measurement of spatial aptitude by dynamic assessment." *Intelligence*, 11, 333–358.

Feuerstein, R. (1979). *The dynamic assessment of retarded performers: The learning potential assessment device, theory, instruments, and techniques*. Baltimore, MD: University Park Press.

Haygood, R.C., & D.F. Johnson. (1983). "Focus shift and individual differences in the Sternberg memory-search task." *Acta Psychologica, 53*, 129–139.

Hunt, E. (1974). "Quote the Raven? Nevermore!" In *Knowledge and Cognition*, edited by L.W. Gregg. Potomac, MD: Lawrence Erlbaum.

Ippel, M.J., & A.L. Beem. (1987). "A theory of antagonistic strategies." In *Learning and Instruction: European Research in an International Context*(Vol. 1), edited by E. De Corte, H. Lodewijks, R. Parmentier, & P. Span. Oxford: Leuven University Press and Pergamon Press.

Jacobs, P.I., & M. Vandeventer. (1972). "Evaluating the teaching of intelligence." *Educational and Psychological Measurement, 32*, 235–248.

Kirby, J.R., & M.J. Lawson. (1983). "Effects of strategy training on Progressive Matrices performance." *Contemporary Educational Psychology, 8*, 127–140.

Lawson, M.J., & J.R. Kirby. (1981). "Training in information processing algorithms." *British Journal of Educational Psychology, 51*, 321–335.

Lawry, J.A., M.C. Welsh, & W.E. Jeffrey. (1983). "Cognitive tempo and complex problem solving." *Child Development, 54*, 912–920.

Malloy, T.E., C. Mitchell, & O.E. Gordon. (1987). "Training cognitive strategies underlying intelligent problem solving." *Perceptual and Motor Skills, 64*, 1039–1046.

Marshalek, B., D.F. Lohman, & R.E. Snow. (1983). "The complexity continuum in the radex and hierarchical models of intelligence." *Intelligence, 7*, 107–127.

Nickerson, R.S. (1988). "On improving thinking through instruction." In *Review of research in education* (Vol. 15), edited by E.Z. Rothkopf. Washington, DC: American Educational Research Association.

Raven, J.C. (1962). *Advanced Progressive Matrices. Sets I and II.* London: H.K. Lewis.

Schmid, J., & J. Leiman. (1957). "The development of hierarchical factor solutions." *Psychometrika, 22*, 53–61.

Sternberg, R.J. (1977). *Intelligence, information processing, and analogical reasoning*. Hillsdale, NJ: Lawrence Erlbaum.

———. (1986). *Intelligence applied: Understanding and increasing your intellectual skills*. San Diego, CA: Harcourt Brace Jovanovich.

Sternberg, S. (1966). "High speed scanning in human memory." *Science, 153*, 652–654.

Turner, R.R., V.C. Hall, & S. Grimmett. (1973). "Effects of familiarization feedback on the performance of lower-class and middle-class kindergartners on the Raven Colored Progressive Matrices." *Journal of Educational Psychology, 65*, 356–363.

Undheim, J.O., & J.E. Gustafsson. (1987). "The hierarchical organization of cognitive abilities: Restoring general intelligence through the use of linear structural relations (LISREL)." *Multivariate Behavioral Research, 22*, 149–171.

Welsh, M.C. (1987). "Verbal mediation underlying inductive reasoning: Cognitive tempo differences." *Cognitive Development, 2*, 37–57.